基于轻量化改进YOLOv5的苹果树产量测定方法

李志军[1,2], 杨圣慧[1,2], 史德帅[1,2], 刘星星[1,2], 郑永军[1,2*]

（1. 中国农业大学 工学院，北京 100083；2. 中国农业大学 烟台研究院，山东烟台 264670）

摘　要：果树测产是果园管理的重要环节之一，为提升苹果果园原位测产的准确性，本研究提出一种包含改进型YOLOv5果实检测算法与产量拟合网络的产量测定方法。利用无人机及树莓派摄像头采集摘袋后不同着色时间的苹果果园原位图像，形成样本数据集；通过更换深度可分离卷积和添加注意力机制模块对YOLOv5算法进行改进，解决网络中存在的特征提取时无注意力偏好问题和参数冗余问题，从而提升检测准确度，降低网络参数带来的计算负担；将图片作为输入得到估测果实数量以及边界框面总积。以上述检测结果作为输入、实际产量作为输出，训练产量拟合网络，得到最终测产模型。测产试验结果表明，改进型YOLOv5果实检测算法可以在提高轻量化程度的同时提升识别准确率，与改进前相比，检测速度最大可提升15.37%，平均*mAP*最高达到96.79%；在不同数据集下的测试结果表明，光照条件、着色时间以及背景有无白布均对算法准确率有一定影响；产量拟合网络可以较好地预测出果树产量，在训练集和测试集的决定系数R^2分别为0.7967和0.7982，均方根误差*RMSE*分别为1.5317和1.4021 kg，不同产量样本的预测精度基本稳定；果树测产模型在背景有白布和无白布的条件下，相对误差范围分别在7%以内和13%以内。本研究提出的基于轻量化改进YOLOv5的果树产量测定方法具有良好的精度和有效性，基本可以满足自然环境下树上苹果的测产要求，为现代果园环境下的智能农业装备提供技术参考。

关键词：苹果原位测产；深度学习；果实检测；BP神经网络；YOLOv5

中图分类号：S252+.9　　**文献标志码：**A　　**文章编号：**202105-SA005

引用格式：李志军, 杨圣慧, 史德帅, 刘星星, 郑永军. 基于轻量化改进YOLOv5的苹果树产量测定方法[J]. 智慧农业(中英文), 2021, 3(2): 100-114.

LI Zhijun, YANG Shenghui, SHI Deshuai, LIU Xingxing, ZHENG Yongjun. Yield estimation method of apple tree based on improved lightweight YOLOv5[J]. Smart Agriculture, 2021, 3(2): 100-114. (in Chinese with English abstract)

1　引　言

果树产量测定不但可以帮助果农掌握果树的生长情况、估算果园的整体产值，而且可以为合理安排收获提供定量依据[1]。传统的果树测产方法主要依靠人工目测清点，不仅对测产人员的经验有较高要求，而且劳动强度大、精度较低[2]。

为实现苹果测产流程的自动化，学者开始利用机器视觉等技术进行相关研究，主要集中在利用该技术从果树图像提取出果实个数等信息[3-5]，而基于图像信息估测果树产量的研究仍需进一步深入。程洪等[6]提出将果实区域比例、果实个数比、小面积果实比例、果实树叶比等作为特征输入，建立神经网络的方法拟合果树产量。

收稿日期：2021-05-13　修订日期：2021-06-10

基金项目：山东烟台校地融合发展项目（2019XDRHXMPT30）；国家重点研发项目（2018YFD0700603）

作者简介：李志军（1996—），男，硕士研究生，研究方向为智能农业装备。E-mail：335022969@qq.com。

*通讯作者：郑永军（1973—）男，博士，教授，研究方向为智能农业装备。电话：13810868016。E-mail：zyj@cau.edu.cn。

Crtomir 等[7]从果树捡果结束到果实收获期内，采集“Golden Delicious”和“Braeburn”两种苹果树图像数据，以果实个数为输入，产量为输出构建人工神经网络进行模型的训练和测试。该方法需要进行多组数据的采集，因此仅适用于接近或已经位于成熟期的苹果果树测产。Roy 等[8]提出了一种基于颜色识别苹果的半监督聚类方法，以及一种利用空间属性从具有任意复杂几何形状的苹果簇中估计数量的无监督聚类方法，将其集成为一个完整的端到端计算机视觉系统，使用单个摄像机捕获的图像作为输入，输出果园的预测产量，在不同数据集上的准确度为91.98%～94.81%。

基于深度学习的目标检测算法可以快速检测出目标数量，其主要分为两类，一类是以YOLO（You Only Look Once）系列[9-12]和SSD（Single Shot MultiBox Detector）系列[13-15]为代表的一阶段检测算法，该类算法具有较快的检测速度，但是精度相对偏低；另一类是以区域卷积神经网络（Region-CNN，R-CNN）系列[16-18]为代表的二阶段检测算法，其检测精度较高，缺点是实时性差。YOLOv5凭借其较快的检测速度以及良好的检测精度得到了研究人员的青睐[19,20]。具体来看，YOLOv5在数据的输入端增加了Mosaic数据增强、自适应锚框计算、自适应图片缩放等操作；特征提取网络为基于CSPNet[21]的CSPDarknet53，可以在一定范围内降低内存损耗；处理输出部分采用FPN[22]和PANet[23]结构，可以加快各层级之间的信息流通。

本研究以自然环境下的苹果为研究对象，通过对YOLOv5检测算法进行轻量化改进，考虑摘袋后不同着色时间、不同光照条件以及背景有无白布对结果的影响进行数据分析；融合产量拟合网络，建立苹果树测产模型，将图像数据作为输入，估测苹果树产量，为果实收获期合理安排采摘人员提供参考，为现代果园环境下的智能农业装备提供技术参考。

2 数据采集与处理

2.1 图像数据采集

原位图像在山东省烟台市栖霞市官道镇姚庄村山东通达现代农业集团有限公司果园基地（北纬37°16′，东经120°64′）采集，苹果品种为“烟富3号”，摘袋后着色16～22 d采摘果实。采集设备为自制的四旋翼无人机（图1），搭载树莓派4B作为图像采集与存储核心，其CPU为Cortex-A72@1.5GHz，GPU为Broadcom VideoCore VI，运行内存8 GB，存储容量128 G。摄像头（Raspberry Pi Camera V2），像素为500万，采集频率为30 Hz，影像最大光圈为F2.35，焦距为3.15 mm，视场角为65°。

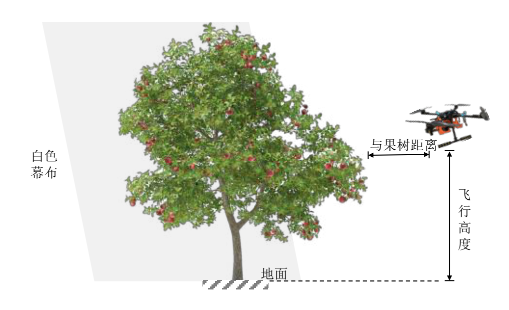




（a）本研究搭建的采集平台



（b）果园原位图像采集现场



（c）图像采集过程示意图

图1 苹果果园原位图像采集

Fig.1 Acquisition of in-situ images of the apple trees in orchard

为降低其他果树产生的干扰，采用长4 m、高3 m的白色幕布为背景，跟随无人机移动（如图1（c）所示）。图像采集时无人机飞行高度为1.5 m，距离果树1.2 m，在晴天时采集顺光、侧光、逆光三个角度的有白色背景与自然条件图像，作为测产模型的训练及在自然环境下的应用

效果验证。

着色 1 d的苹果颜色呈现青绿色，与果树叶片颜色较为接近；8 d时开始上色，颜色呈现淡红色；着色15 d即可完全上色。不同着色时间的数据，便于对本研究提出的果实检测算法进行泛化能力测试。图像采集日期为2020年10月3日至17日，每隔7 d采集一次图像，采集时间为上午10点至下午4点，共采集到苹果着色1 d、着色8 d、着色15 d的三组数据，在果实着色16 d时采集果树产量数据（图2）。

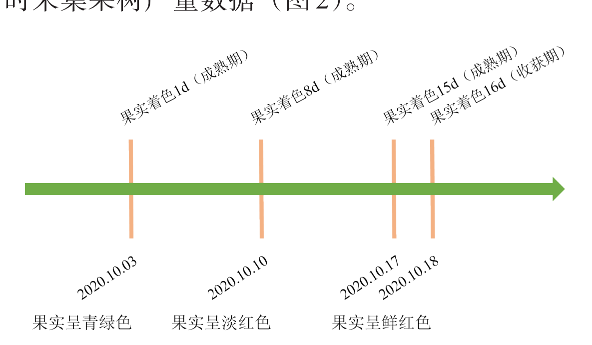


图2　苹果树产量测定数据采集时间轴

Fig. 2 Data collection timeline of the apple trees yield estimation

2.2 图像数据预处理

2.2.1 数据清洗

为降低重复图片数量以及无果实图片对模型训练的干扰，使用人工筛选的方法对采集图像进行数据清洗，即删除因无人机悬停造成的重复图片以及无人机姿态调整过程中不含苹果的图片。数据清洗后，着色1 d、8 d和15 d部分数据如图3所示。利用不同着色时间采集的不同光照条件数据，分析不同着色时间以及不同光照条件数据对检测算法的影响。图像数据包含果树整体图像以及局部图像，检测算法在工作过程中只迭代训练标注过的果实区域。本研究无人机拍摄的整体图像与局部图像果实区域大小相近、分布相似，故将整体与局部图像作为训练数据可以使算法在不损失检测准确率的同时提高其泛化能力。

利用不同时间点采集的背景无白布数据，与背景有白布数据进行对比分析，验证本研究提出检测算法的实际应用效果，背景无白布下不同时间点采集的部分苹果图片如图4所示。

2.2.2 数据集划分与标注

数据清洗后，着色1 d、8 d和15 d数据各保留1000张。从背景有白布数据中随机挑选300张图片作为测试集1，并将测试集1分为顺光、侧光和逆光三个子集，每个子集包含100张图片；从背景无白布数据中挑选300张作为测试集2，不分光照条件子集；剩下的图片均作为训练集数据。

采用人工划分方法，在labelImg软件中对目标果实进行框选，包含目标框的位置坐标、类别等信息，标注结果如图5（第104页）所示。

标注完成后，将样本数据集制作成标准PASCAL VOC2012格式的数据集，图片及目标框数量如表1所示。

2.3 产量数据采集

产量数据于2020年10月18日9:00—17:00采集，此时苹果着色16 d。采集时对单株果树进行

表1　样本数据集及数据量

Table 1 Sample data-sets and data volume

<table>
<tr><th rowspan="3">着色天数/d</th><th colspan="2" rowspan="2">训练集</th><th colspan="6">测试集1</th><th colspan="2" rowspan="2">测试集2</th></tr>
<tr><th colspan="2">顺光子集</th><th colspan="2">侧光子集</th><th colspan="2">逆光子集</th></tr>
<tr><th>图片数量/张</th><th>目标框数量/个</th><th>图片数量/张</th><th>顺光目标框数量/个</th><th>图片数量/张</th><th>侧光目标框数量/个</th><th>图片数量/张</th><th>逆光目标框数量/个</th><th>图片数量/张</th><th>目标框数量/个</th></tr>
<tr><td>1</td><td>600</td><td>15,017</td><td>100</td><td>3282</td><td>100</td><td>3027</td><td>100</td><td>2963</td><td>100</td><td>3125</td></tr>
<tr><td>8</td><td>600</td><td>16,639</td><td>100</td><td>3155</td><td>100</td><td>3241</td><td>100</td><td>2834</td><td>100</td><td>3272</td></tr>
<tr><td>15</td><td>600</td><td>15,892</td><td>100</td><td>3268</td><td>100</td><td>3114</td><td>100</td><td>3052</td><td>100</td><td>3136</td></tr>
<tr><td>总计</td><td>1800</td><td>47,548</td><td>300</td><td>9705</td><td>300</td><td>9382</td><td>300</td><td>8849</td><td>300</td><td>9533</td></tr>
</table>


(a)着色1 d顺光 (b)着色1 d侧光 (c)着色1 d逆光
(d)着色8 d顺光 (e)着色8 d侧光 (f)着色8 d逆光
(g)着色15 d顺光 (h)着色15 d侧光 (i)着色15 d逆光

图3 不同时间点采集的苹果图像(背景有白布)

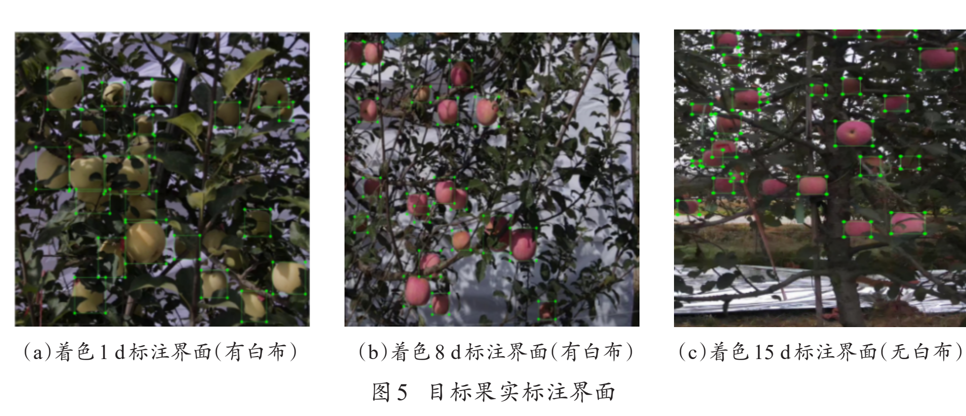
Fig.3 Apple images captured at different time points (with white cloth in background)


(a)着色1 d图像 (b)着色8 d图像 (c)着色15 d图像

图4 不同时间点采集的苹果图像(背景无白布)

Fig. 4 Apple images captured at different time points(without white cloth in background)

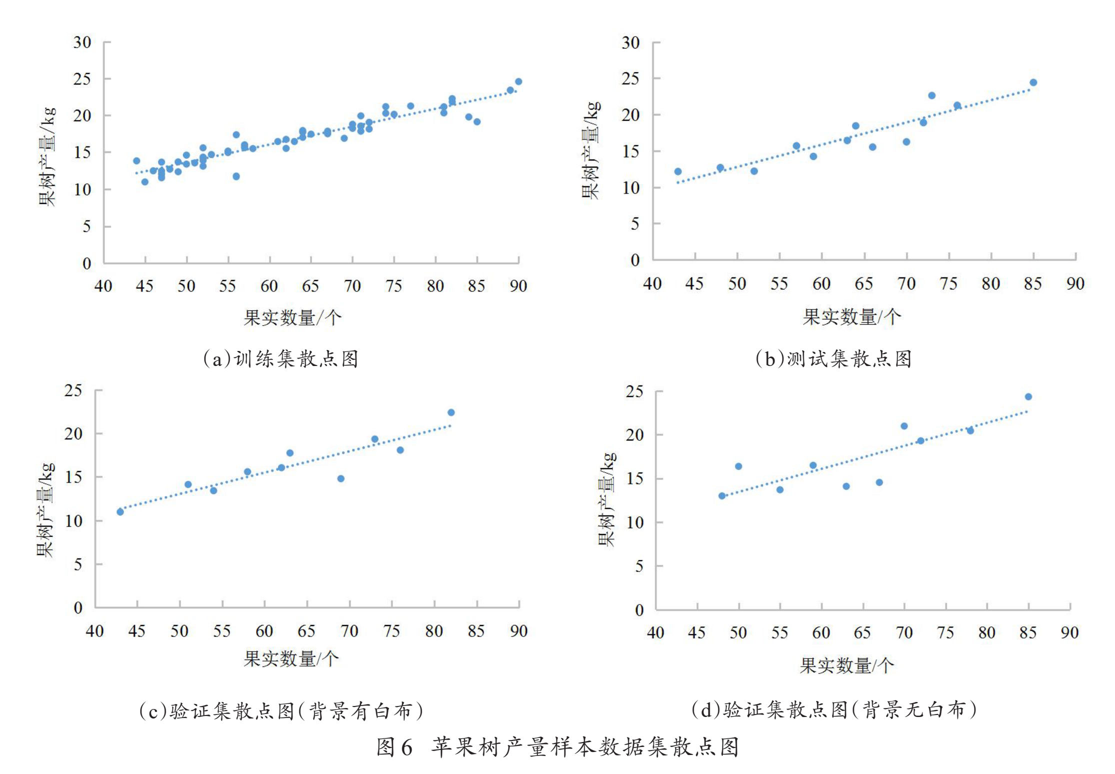
编号，在每株果树收获前，使用无人机拍摄当前果树图像；收获后，将单株果实放置在同一个框中，使用电子秤对苹果进行称重，去掉框的重量即可得到单株果树产量。得到数据共93组，每组数据包含果树图像以及相应产量，其中60组用于产量拟合网络训练，13组用于产量拟合网络测试，10组用于背景有白布的测产模型验证，10组用于背景无白布的测产模型验证。将不同产量数据集数据绘制成散点图，如图6所示。果实数量与果树产量之间呈现一定的线性相关性，并且果实平均重量在250~280 g之间，表明该果园的果树长势较好，果树个体之间无明显差异。



(a)着色1 d标注界面(有白布)　(b)着色8 d标注界面(有白布)　(c)着色15 d标注界面(无白布)

图5　目标果实标注界面

Fig. 5 Tagging interface for target fruit



(a)训练集散点图　(b)测试集散点图

(c)验证集散点图(背景有白布)　(d)验证集散点图(背景无白布)

图6　苹果树产量样本数据集散点图

Fig. 6 Scatter plot of yield sample data sets of the apple trees

3　苹果树测产模型

3.1 模型总体结构

本研究提出的果树测产模型分为果实检测算法和产量拟合网络两部分，如图7所示。果实检测算法利用改进型YOLOv5对输入的果树图像进行目标检测，输出图像中果实数量以及所有果实边界框总面积；产量拟合网络负责将上述算法的输出作为当前网络的输入，利用BP神经网络拟合果树产量。检测算法部分使用图像训练集、拟合网络部分使用产量训练集分开进行训练，训练完成后，可实现在测产模型中输入果树图像即可直接输出相应的果树产量。

3.2 改进型YOLOv5果实检测算法

3.2.1　基于轻量化改进的YOLOv5

YOLOv5模型起源于YOLO，该算法在输出层回归目标框的位置坐标及其所属类别具有良好的检测速度。YOLO算法的核心思想是将输入图片划分为7×7个网格，目标中心所在的网格负责预测该目标。每个网格负责预测2个目标框，

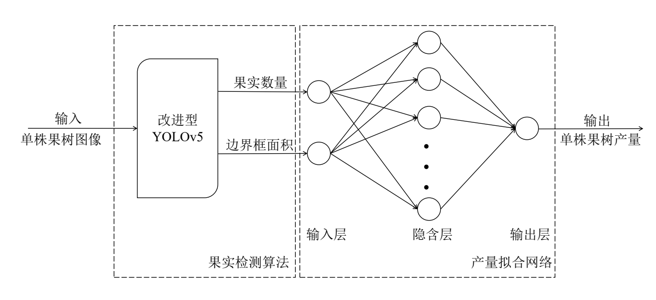


图7　模型总体结构图

Fig. 7 Overall structure of the model

该目标框回归位置坐标以及预测置信度值[24]。设定一个置信度阈值，滤除置信度较低的目标框，并对保留的框进行非极大值抑制（Non-Maximum Suppression，NMS）处理，得到最终的预测效果，如图8所示。

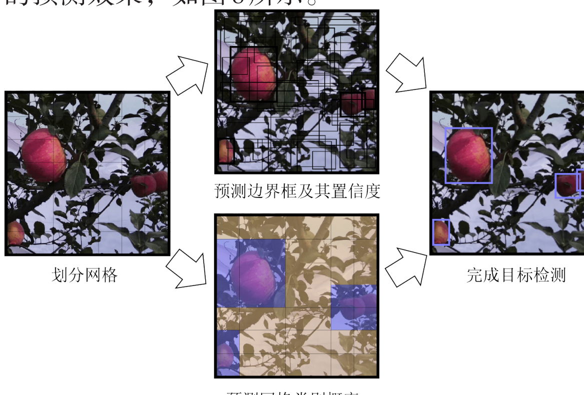


图8　YOLO目标检测模型

Fig. 8 YOLO object detection model

然而，传统的YOLOv5网络参数量较大，且在提取过程中存在无注意力偏好问题，即对不同重要程度的特征采用相同的加权方式。因此，本研究提出将YOLOv5特征提取网络中的标准卷积更换为轻量化的深度可分离卷积，并且基于深度可分离卷积[25]和视觉注意力机制提出一个池化注意力模块（Pooling Block Attention Module，PBAM），将该模块添加到YOLOv5网络中解决无注意力偏好问题。PBAM利用先压缩后扩张的方式，将浅层特征采样到的关键点进行增强学习，并且该模块引入了类似残差单元的结构，保证了网络在较深的情况下不会出现梯度消失或是梯度爆炸的问题；该模块输出特征图与输入特征图分辨率保持一致，在不更改网络结构的前提下，PBAM可以嵌入到任意网络结构中，具有结构简单、使用方便等优势，通过建立通道间的相互依赖关系，从而达到自适应校准通道间相应特征的目的。图9为改进后的YOLOv5算法框图，其中，红框为本研究改进部分。

融合后的YOLOv5算法不仅可以利用深度可分离卷积减少注意力机制模块带来的计算压力，而且可以将卷积计算得到的浅层特征图抽取为重要信息更多的深层特征图，进一步提取关键信息，提升算法的整体检测效果。

3.2.2　损失函数计算

YOLOv5的损失函数由边界框置信度损失（L_{conf}）、类别损失（L_{cla}）以及坐标损失（L_{GIOU}）三部分构成，改进后的YOLOv5算法仅深化了网络深度，对以上函数并无影响，不需要构建新的损失函数。其中置信度损失和类别损失采用交叉熵的方法进行计算，如公式（1）～（3）所示。

$$L_{total} = L_{conf} + L_{cla} + L_{GIOU} \tag{1}$$

$$L_{conf} = \lambda_{obj} \sum_{i=0}^{S2} \sum_{j=0}^{B} I_{ij}^{obj} [-\hat{C}_i \ln C_i - (1 - \hat{C}_i) \ln (1 - C_i)] + \lambda_{nobj} \sum_{i=0}^{S2} \sum_{j=0}^{B} I_{ij}^{nobj} [-\hat{C}_i \ln C_i - (1 - \hat{C}_i) \ln (1 - C_i)] \tag{2}$$

$$L_{cla} = \sum_{i=0}^{S2} \sum_{j=0}^{B} \sum_{c \subset cla} I_{ij}^{obj} [-\hat{p}_i(c) \ln (p_i(c)) - (1 - \hat{p}_i(c)) \ln (1 - p_i(c))] \tag{3}$$

其中，S^2为划分的网格数量，个；B为每个网格预测边界框数量，个；I_{ij}^{obj}为判断第i个网格的第j个边界框是否有需要预测的目标；I_{ij}^{nobj}为判断第i个网格的第j个边界框是否有不需要预测的目标；λ_{obj}和λ_{nobj}为网格有无目标的权重系数；$C_i, \hat{C}_i$为预测目标和实际目标的置信度值；c为边界框预测的目标类别；$p_i(c)$为第i个网格检测到目标时，其所属c的预测概率，%；$\hat{p}_i(c)$为第i

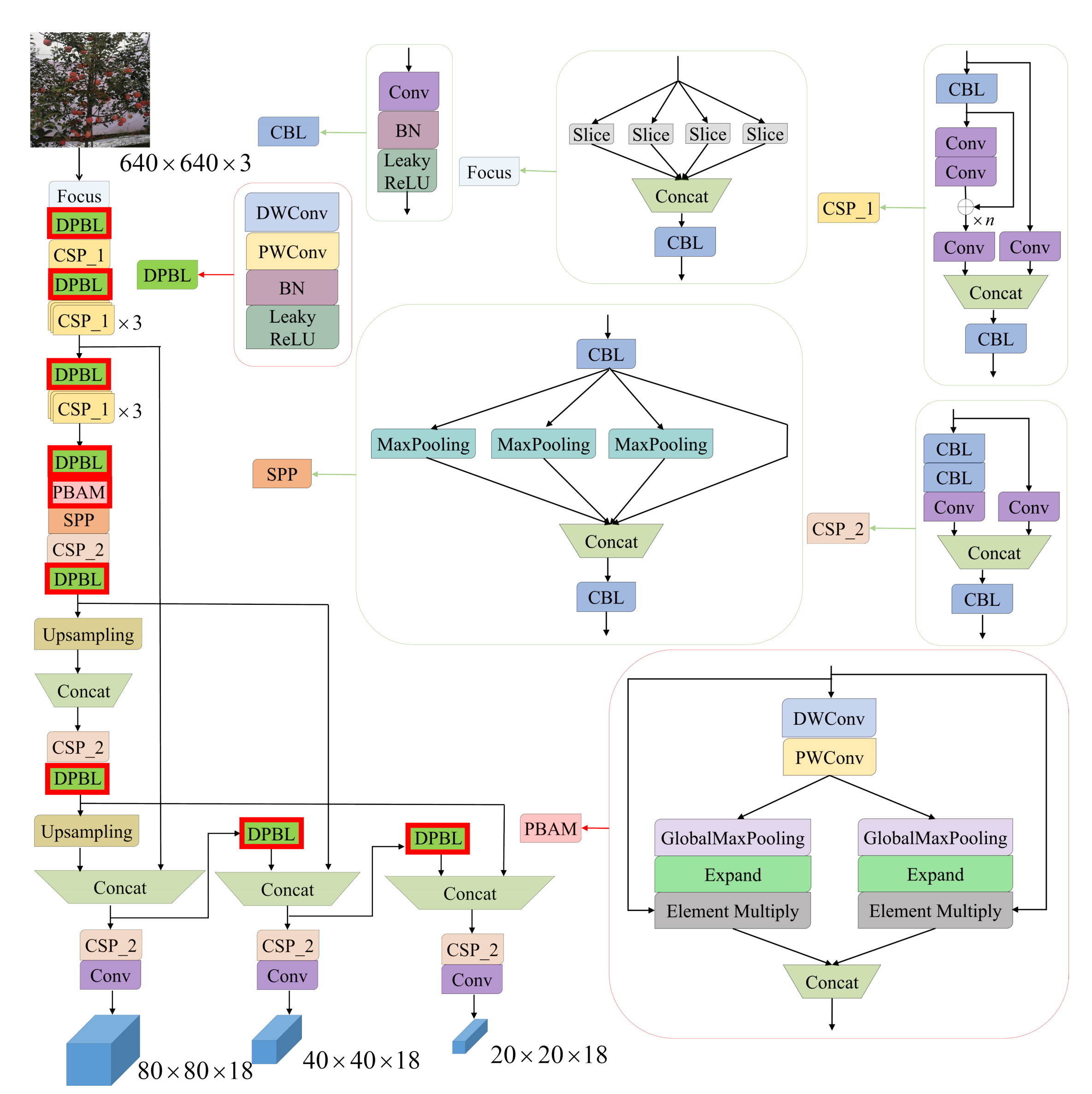


图9　改进后的YOLOv5算法框图

Fig. 9　Block diagram of the improved YOLOv5 algorithm

个网格检测到目标时，其所属c的实际概率，%。

本研究采用L_{GIoU}作为边界框坐标的损失函数，如公式（4）～（6）所示。

$$IoU = \frac{A \cap B}{A \cup B} \tag{4}$$

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|} \tag{5}$$

$$L_{GIoU} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} (1 - GIoU) \tag{6}$$

其中，A为真实框的面积，像素；B为预测框的面积，像素；C为A和B的最小外接矩形面积，像素。

3.3 产量拟合网络

由于果实检测数量以及边界框面积与估测产量之间的函数映射关系并不明确，属非线性函数关系特征，因此本研究采用BP神经网络作为果实产量拟合部分，提升模型整体对任意复杂模式的分类能力与多维函数映射能力。由于BP网络的输入包含了边界框面积，因此图像拍摄的距离

与相机的参数就显得尤为重要。本研究在图像采集时，保持采集平台与果树的距离不变，并且固定相机的参数，保证测产的准确性。

本研究采用3层全连接层、1层ReLU激活层和1层Sigmoid激活层完成网络结构的搭建，网络拓扑结构如图10所示。BP网络的输入神经元数量为2个，分别对应改进型YOLOv5果实检测算法输出的果实数量（个）以及边界框总面积（像素）；输出神经元的数量为1个，对应图像中的果树产量。在隐藏层中加入ReLU或Sigmoid激活函数，可以在一定程度上增加神经网络的非线性因素，加快算法训练速度，并且可以解决反向传播时梯度消失问题，有效降低过拟合发生的概率。

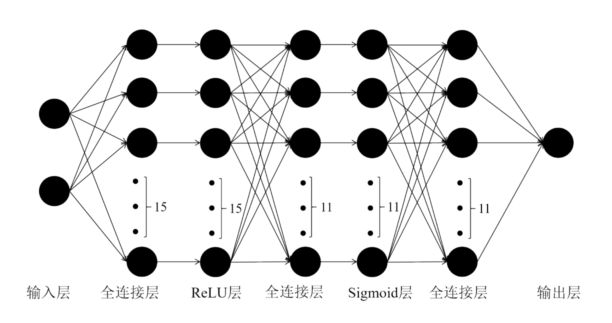


图10 BP网络拓扑结构图

Fig.10 BP network topology structure diagram

隐藏层神经元数量的确定没有明确的理论方法，因此本研究先由经验公式确定初始值，如公式（7）所示；再根据网络训练过程中的误差表现选取最优值。最终的隐藏层神经元数量为15个或11个，如图10所示。

$$N = \sqrt{m + n} + \mathrm{a} \tag{7}$$

其中，N为隐藏层节点数量，个；m为输入神经元数量，个；n为输出神经元数量，个；a为1～10之间的常数。

利用BP网络进行数据拟合的主要步骤如下。

（1）数据归一化。为保证性能的稳定性，分别对果实数量、边界框面积以及果树产量进行输入样本的归一化处理。通过除以归一化系数，将输入特征和输出产量归一到0～1之间。

（2）BP网络训练。可分为四步，首先，初始化网络权重；其次，数据正向传播；再次，误差反向传播；最后，网络权重与神经网络元偏置调整。

（3）数据反归一化。为获取果树产量对应的常量，需要对预测值进行反归一化。将预测数据乘以对应的归一化系数，将该数据重新映射到原始区间，得到最终产量。

3.4 模型训练

模型训练分为两个阶段，第一阶段为果实检测，通过训练目标检测算法，预测图片中果实数量及边界框总面积；第二阶段为产量预测，基于BP神经网络拟合数据集的果实数量、边界框面积和产量。

采用PyTorch深度学习框架进行模型的搭建，系统硬件配置为AMD Ryzen7 4800H CPU@2.9GHz处理器，6 GB NVIDIA GeForce GTX 1660Ti GPU，16 GB运行内存，512 GB SSD硬盘容量，训练和测试所用的操作系统为Windows10，64位系统。代码编译器为PyCharm2019.3.3社区版，并且配置了CUDA 10.2和cuDNN 7.6.5进行GPU加速工作。

4 算法测试与试验

4.1 果实检测算法性能分析

本研究选取平均准确率（mAP）作为算法的整体评价指标。查准率（P）是指被预测为正例的样本中实际为正样本的比例，查全率（R）是指实际为正例的样本中被预测为正样本的比例，根据查准率和查全率之间的关系可以绘制查准率—查全率曲线（P-R曲线）。所有类别的准确率（AP）是指曲线与坐标轴围成区域的面积，求出所有类别的AP并取均值，可得各类别的mAP，计算如公式（8）～（11）所示。

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$AP = \frac{1}{N}\sum p(r) \quad (10)$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (11)$$

其中，*TP* 为正样本预测正确的数量，个；*FN* 为负样本预测错误的数量，个；*FP* 为正样本预测错误的数量，个；*TN* 为负样本预测正确的数量，个；$p(r)$ 为不同查准率 r 下对应的查全率 p；AP_i 为第 i 类的检测准确率；N 为类别数量，个。

为验证本研究提出的改进型YOLOv5果实检测算法的可行性，需要对特征提取网络中单独更换深度可分离卷积、单独嵌入注意力机制模块、融合深度可分离卷积和注意力机制模块的YOLOv5算法与未改进的YOLOv5算法进行检测速度、平均准确率性能测试，性能测试完成后进行背景有白布以及背景无白布数据测试。

（1）检测速度算法测试。本研究对算法检测多张图片的速度进行记录，并取其均值，对比试验结果如表2所示，包括以改进前算法检测速度为基础不同改进版本算法的相对提升率。其中YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x 分别是YOLOv5的4个不同深度和宽度的网络结构，其参数量依次递增。

表2 改进前后算法检测速度的对比试验结果

Table 2 Comparison experiment results of the detection speed of the algorithm before and after the improvement

算法		检测速度及相对提升率	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
改进前		检测速度/ms	37.16	84.87	152.33	310.53
		相对提升率/%	0.00	0.00	0.00	0.00
改进后	单独更换深度可分离卷积	检测速度/ms	30.73	70.85	132.26	275.54
		相对提升率/%	−17.28	−16.51	13.17	−11.26
	单独嵌入注意力机制模块	检测速度/ms	37.83	87.24	155.87	318.66
		相对提升率/%	1.80	2.79	2.32	2.61
	融合深度可分离卷积和注意力机制模块	检测速度/ms	31.44	72.90	136.28	281.49
		相对提升率/%	−15.37	−14.10	−10.50	−9.35

单独更换深度可分离卷积，使算法的检测速度明显提升，提升幅度为11.26%～17.28%。单独嵌入注意力机制模块会对计算速度有一定影响，但是影响较小，与改进前YOLOv5算法检测速度基本持平。融合深度可分离卷积和注意力机制模块后，与单独更换深度可分离卷积相比，算法会增加部分计算负担，降低部分检测速度；检测速度最快为31.44 ms，与改进前相比提升最大15.37%。上述结果表明，单独更换深度可分离卷积带来计算速度的提升较大；融合注意力机制模块后，与改进前的算法相比仍具有较快的检测速度优势。

（2）平均准确率算法测试。本研究对算法检测多张图片的 *mAP* 进行记录，对比试验结果如表3所示，包括以改进前算法平均准确率为基础不同改进版本算法的绝对提升率。

改进前的YOLOv5算法在不同结构下的 *mAP* 分别是89.83%、90.75%、92.07%和93.44%，除了改进后单独更换深度可分离卷积YOLOv5s的检测准确率低于90%，其余结构均在90%以上，该结果表明YOLOv5拥有较好的模型结构以及检测效果，对自然环境下的苹果检测具有较好的适用性。单独更换深度可分离卷积会使算法损失少量准确率，与改进前相比，损失率基本控制在1%以内。不同结构下算法最高准确率与最低准确率最大相差4.35%，出现在单独嵌入注意力机制模块的YOLOv5算法检测中，在该种改进方式下，检测准确率与改进前相比最大可提升3.56%，平均提升3.17%。通过在算法结构中嵌入注意力机制模块，有助于将卷积计算得到的浅

表3 改进前后算法*mAP*对比试验结果

Table 3 Comparison of experiment results of the original algorithm and the improved algorithm of the *mAP*

单位:%

算法		平均准确率及绝对提升率	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
改进前		平均准确率	89.83	90.75	92.07	93.44
		绝对提升率	0.00	0.00	0.00	0.00
改进后	单独更换深度可分离卷积	平均准确率	88.94	90.05	91.55	92.77
		绝对提升率	−0.89	−0.70	−0.52	−0.67
	单独嵌入注意力机制模块	平均准确率	92.10	94.62	95.63	96.45
		绝对提升率	2.27	3.87	3.56	3.01
	融合深度可分离卷积和注意力机制模块	平均准确率	92.88	93.99	95.15	96.27
		绝对提升率	3.05	3.24	3.08	2.83

层特征图抽取为重要信息更多的深层特征图，学习通道间的相关性，把特征图中重要的特征增强，次要的特征减弱，对关键信息进行更进一步的提取，有效提升算法的整体检测效果。融合后的YOLOv5l和YOLOv5x检测准确率均超过95%，其中YOLOv5x的检测准确率达到96.27%，为测试结果最高。以上结果均表明改进后的YOLOv5算法具有较高的检测准确率。

（3）背景有白布算法测试。在自然光照条件下，太阳光作为主要的光源，不同的拍摄角度会在一定程度上影响目标检测的*mAP*。本研究选取了不同时间点下不同光照角度的苹果图像作为样本数据集，对改进后的YOLOv5模型进行迭代训练，目标检测效果如图11。

顺光拍摄的照片明亮清晰，不会出现明显的光影变化；侧光拍摄的照片层次分明，物体有较为明显的轮廓；逆光拍摄的照片阴暗模糊，容易出现曝光不足。本研究针对上述不同数据集，并且与YOLOv5、YOLOv3、SSD等模型进行对比实验，各模型检测结果如表4所示。

改进的YOLOv5在各种不同的光源和时间点的数据中，全部得到了最好的检测结果。最高*mAP*出现在着色15 d侧光数据集，达到96.79%，在不同数据集中表现均优于YOLOv5、YOLOv3、SSD，在不同测试集中平均*mAP*为93.30%，说明改进YOLOv5算法具有较好的检测性能；所有算法在侧光条件下检测效果最好，在逆光条件下检测效果最差的原因是：逆光下树叶与果实的颜色较暗，果实边缘不够清晰，容易造成混淆，增加目标检测难度；算法效果随着色时间提高，原因在于着色1 d果实颜色与树叶颜色较为接近，呈现出绿色，容易误识别，而第15 d果实颜色较为鲜艳，可以明显与周围叶片区分开，检测准确率较高。

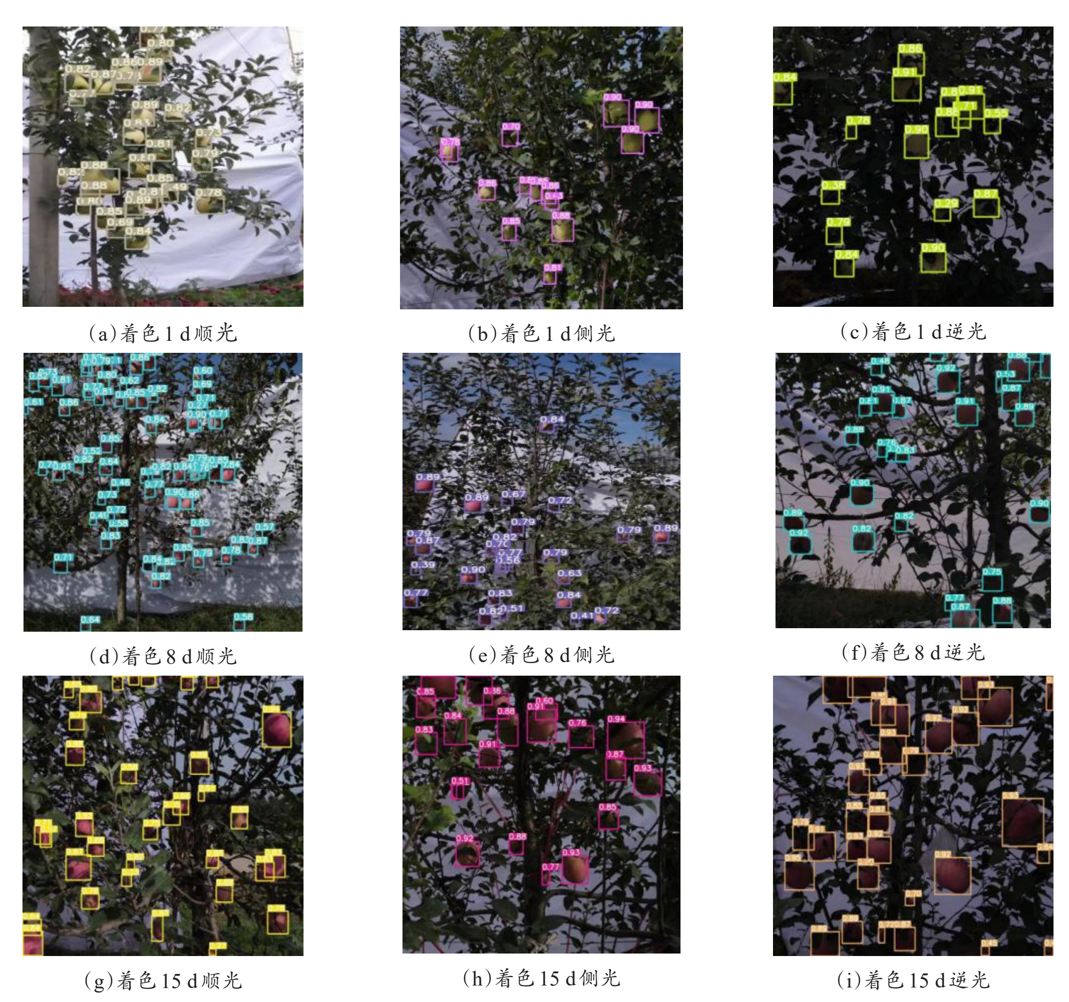
（4）背景无白布算法测试。为验证改进型YOLOv5果实检测算法在自然环境下的应用效

表4 不同数据集的*mAP*对比试验结果

Table 4 *MAP* comparison results of different data sets

单位:%

模型	着色1 d数据集			着色8 d数据集			着色15 d数据集		
	顺光	侧光	逆光	顺光	侧光	逆光	顺光	侧光	逆光
改进型YOLOv5	90.25	92.48	89.31	93.56	95.08	92.95	95.26	96.79	94.07
YOLOv5	89.56	90.57	87.35	91.78	92.23	90.57	93.57	94.29	92.38
YOLOv3	87.82	88.11	86.77	89.93	90.86	88.75	92.27	93.33	91.43
SSD	86.26	86.98	83.67	87.69	89.54	88.37	90.25	91.33	89.40



(a)着色1 d顺光　(b)着色1 d侧光　(c)着色1 d逆光

(d)着色8 d顺光　(e)着色8 d侧光　(f)着色8 d逆光

(g)着色15 d顺光　(h)着色15 d侧光　(i)着色15 d逆光

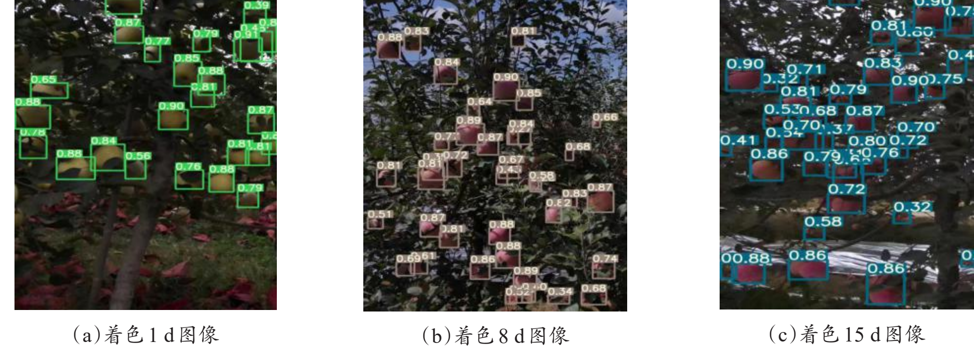
图11　不同时间点不同光照角度的果实检测效果(背景有白布)

Fig. 11 Detection results under different illumination angles at different time instances(with white cloth in background)

果，对无白色背景的着色1 d、8 d和15 d苹果图像进行测试，结果如图12。

结果可知，算法均可完成检测任务，对近处的果实识别率较高。由于近处的苹果区域面积较大，呈现特征较多，算法对其预测的置信度较高；远处的苹果在图像上占据的像素点数量较少，呈现的特征较少，识别置信度受到影响。

表5为无白布背景下的算法 *mAP* 对比结果，



(a)着色1 d图像　(b)着色8 d图像　(c)着色15 d图像

图12　不同时间点背景无白布的果实检测效果

Fig. 12 Fruit detection results without white cloth in background at different times in stances

与背景有白布的算法检测准确率相比，在不同着色时间的数据集下均有一定的准确度损失。但是对于果树测产而言，需要测定的是距离摄像头较近的果树产量，利用算法过滤背景中的果树果实，减少背景果实对其产量的影响，因此在该背景下并不是检测准确率越高越好。本研究的果实检测算法恰能满足识别近处苹果、滤除远处苹果的要求，可适用于背景无白布的果实检测。

表5　背景无白布下的算法*mAP*对比试验结果

Table 5　Comparison experimental results of the *mAP* of the algorithm under the condition of without white cloth in background

单位：%

数据集	YOLOv5s	YOLOv5m	YOLOv5l	YOLOv5x
着色1 d	81.47	83.26	84.63	85.89
着色8 d	83.54	86.72	87.39	88.17
着色15 d	84.76	87.93	89.56	90.72

4.2 产量拟合网络性能分析

采用表6中的相关参数评估产量拟合网络的性能。相关系数R和决定系数R^2用于衡量预测产量和实际产量之间的相关程度，数值越大表明二者的相关性越好；均方根误差*RMSE*用来衡量预测产量和实际产量之间的误差，其值越小表明精确度越高；平均绝对误差*MAE*和平均绝对百分比误差*MAPE*能够反映预测产量偏离实际产量的程度，数值越小，表明二者差别越小，拟合效果越好。在训练集和测试集上预测产量和实际产量的R分别为0.8979和0.8864，R^2分别为0.7967和0.7982，表明果实数量、边界框面积和产量之间线性相关程度较高，曲线拟合较好。

表6　BP估测模块的性能参数

Table 6　The performances of BP yield measurement module

参数	训练集	测试集
R	0.8979	0.8864
R^2	0.7967	0.7982
RMSE/kg	1.5317	1.4021
MAE/kg	1.1259	1.0253
MAPE/%	6.3372	6.2524

对于训练集样本，*RMSE*为1.5317 kg，*MAE*为1.1259 kg，*MAPE*为6.3372%；对于测试集样本，*RMSE*为1.4021 kg，*MAE*为1.0253 kg，*MAPE*为6.2524%。

基于样本训练集建立的产量拟合网络，在测试集上的对比结果如图13所示。可见，该模型可以较好地预测出果树产量，对于不同产量样本的预测精度基本稳定，具有较好的鲁棒性。通过测试结果可知该模型可适用于自然环境下果树收获前的产量测定。

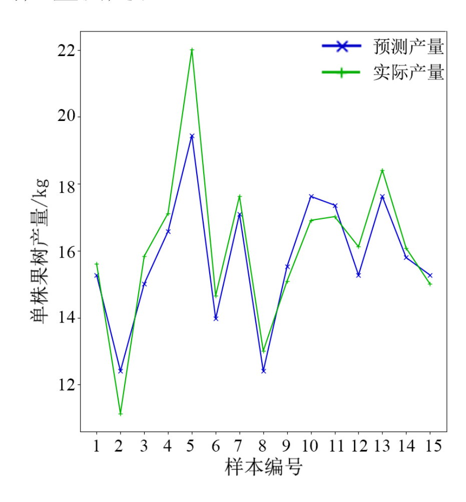


图13　测试集下的预测产量和实际产量的对比

Fig. 13　Comparison of predicted and actual yields under the test data-sets

4.3 测产模型测试

（1）背景有白布的测产模型性能测试。用本研究提出的苹果树测产模型，以产量验证集中背景有白布的果树图像作为输入，输出该株果树的预测产量。由表7所示的相对误差可见，测产模型相对误差的绝对值取值范围为3.05%～6.13%。在对10组数据的预测过程中，仅有第7组的误差稍大，总体相对误差范围基本在7%以内。表明提出的果树测产模型对于输入背景有白布的果树图像，有着较好的产量预测效果；可以通过果实检测算法和产量拟合网络的协调工作，学习到图像数据以及产量数据的重要特征，输出相应的预测产量。

（2）无白布背景的测产模型性能测试。为验

证本研究提出的测产模型在自然环境下的应用效果，选用产量验证集中图像背景无白布的数据集作为模型输入，输出该株果树的预测产量，经过计算后得到表8所示的相对误差。

表7 背景有白布的测产模型相对误差

Table 7 Relative error of yield measurement model with white cloth in background

评价指标	第一组	第二组	第三组	第四组	第五组	第六组	第七组	第八组	第九组	第十组
相对误差δ/%	3.05	4.25	−4.17	3.75	−5.81	4.10	−6.13	−5.39	5.41	4.28

表8 背景无白布的测产模型相对误差

Table 8 Relative error of yield measurement model without white cloth in background

评价指标	第一组	第二组	第三组	第四组	第五组	第六组	第七组	第八组	第九组	第十组
相对误差δ/%	−10.34	9.72	12.15	9.83	11.13	−10.54	−12.71	8.28	11.67	−9.37

与背景有白布的测试结果相比，背景无白布的测产模型相对误差偏大，绝对值的取值范围为8.28%～12.71%。出现这一现象的主要原因在于背景中有其他果树果实的干扰，产量拟合网络在预测产量时会将检测到的果实均判定为当前果树果实，导致结果出现一定的偏差。但是，背景果树中仅有部分特征较多的苹果会被识别，大部分苹果由于距离较远，在图像中占据的像素点较少，检测算法不能识别，因此造成的影响不大，相对误差总体范围在13%以内。上述结果表明，本研究提出的测产模型具有良好的精度和有效性，并且在不同背景下的测产鲁棒性较好，可适用于自然环境下的苹果树产量测定。

5 结 论

本研究提出了一种苹果树测产模型，融合产量拟合网络和改进型YOLOv5果实检测算法，结合数据集预处理，对模型进行了训练和应用，结论如下。

（1）通过更换深度可分离卷积和添加注意力机制模块改进的YOLOv5苹果检测网络，解决网络中存在的特征提取时无注意力偏好问题和参数冗余问题。以图像数据集为输入，得到估测果实数量以及边界框面总积。测试结果表明，该算法可以在提高轻量化程度的同时提升准确率，与改进前相比，检测速度最大可提升15.37%，平均*mAP*最高达到96.79%，在不同数据集下的测试结果表明光照条件、着色时间以及背景有无白布均对算法准确率有一定影响。

（2）以估测果实数量与边界框面积为输入、实际产量为输出，训练产量拟合网络，测试结果表明，产量拟合网络训练集和测试集的R^2分别为0.7967和0.7982，*RMSE*分别为1.5317和1.4021 kg，测产误差较小。

（3）将果实检测算法和产量拟合网络融合得到最终的测产模型。试验结果表明，果树测产模型在背景有白布和无白布的条件下，相对误差范围分别在7%以内和13%以内，证明本研究建立的苹果园原位测产模型具有较好的精度和鲁棒性。如增加更多的样本作为数据输入，可进一步提高目标的识别度和测产的准确性。

参考文献：

[1] 黑龙江省佳木斯农业学校, 江苏省苏州农业学校. 果树栽培学总论[M]. 北京: 中国农业出版社, 2009.
Jiamusi Agricultural School of Heilongjiang province, Suzhou Agricultural School of Jiangsu province. General introduction to fruit cultivation[M]. Beijing: China Agriculture Press, 2009.

[2] 王少敏, 张毅, 高华军, 等. 苹果套袋栽培技术[M]. 济南: 山东科学技术出版社, 2006.
WANG S, ZHANG Y, GAO H, et al. Apple bagging cultivation technology[M]. Jinan: Shandong Science and Technology Press, 2006.

[3] PAPAGEORGIOU E I, AGGELOPOULOU K D, GEMTOS T A, et al. Yield prediction in apples using fuzzy cognitive map learning approach[J]. Computers and Electronics in Agriculture, 2013, 91: 19-29.

[4] AGGELOPOULOU A D, BOCHTIS D, FOUNTAS S,

et al. Yield prediction in apple orchards based on image processing[J]. Precision Agriculture, 2011, 12(3): 448-456.

[5] ZHOU R, DAMEROW L, SUN Y, et al. Using colour features of cv.'Gala' apple fruits in an orchard in image processing to predict yield[J]. Precision Agriculture, 2012, 13(5): 568-580.

[6] 程洪, DAMEROW L, BLANKE M, 等 . 基于树冠图像特征的苹果园神经网络估产模型[J]. 农业机械学报, 2015, 46(1): 14-19.

CHENG H, DAMEROW L, BLANKE M, et al. Ann model for apple yield estimation based on feature of tree image[J]. Transactions of the CSAM, 2015, 46(1): 14-19.

[7] CRTOMIR R, CVELBAR U, TOJNKO S, et al. Application of neural networks and image visualization for early predicted of apple yield[J]. Erwerbs-Obstbau, 2012, 54(2): 69-76.

[8] ROY P, KISLAY A, PLONSKI P, et al. Vision-based preharvest yield mapping for apple orchards[J]. Computers and Electronics in Agriculture, 2019, 164: ID 104897.

[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time objectdetection[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2016: 779-788.

[10] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2017: 7263-7271.

[11] REDMON J, FARHADI A. YOLO v3: An incremental improvement[EB/OL]. 2018. arXiv: 1804.02767v1.

[12] BOCHKOVSKIY A, WANG C, LIAO H. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. 2020. arXiv: 2004.10934.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]// European Conference on Computer Vision. Cham, Switzerland: Springer, 2016: 21-37.

[14] ZHANG S, WEN L, BIAN X, et al. Single-shot refinement neural network for object detection[C]// The IEEE Conference on Computer Vision and Pattern Reco-gnition. Piscataway, New York, USA: IEEE, 2018: 4203-4212.

[15] WANG D, ZHANG B, CAO Y, et al. SFSSD: Shallow feature fusion single shot multibox detector[C]// International Conference in Communications, Signal Processing, and Systems. Cham, Switzerland: Springer, 2019: 2590-2598.

[16] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2014: 580-587.

[17] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway, New York, USA: IEEE, 2015: 1440-1448.

[18] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[19] 周伟鸿, 朱思霖 . 基于深度学习技术的智慧考场方案的应用探究[J]. 信息技术与信息化, 2020(12): 224-227.

ZHOU W, ZHU S. Research on the application of smart examination room solutions based on deep learning technology[J]. Information Technology and Informatization, 2020(12): 224-227.

[20] 王沣 . 改进 YOLOv5 的口罩和安全帽佩戴人工智能检测识别算法[J]. 建筑与预算, 2020(11): 67-69.

WANG F. Artificial intelligence detection and recognition algorithm for masks and helmets based on improved YOLOv5[J]. Construction and Budget, 2020 (11): 67-69.

[21] WANG C Y, LIAO H Y M, YEH I H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]// The IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, WA, USA: CVPRW, 2020: 390-391.

[22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detect-ion[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2017: 2117-2125.

[23] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, New York, USA: IEEE, 2018: 8759-8768.

[24] 李德鑫, 闫志刚, 孙久运 . 基于无人机视觉的河道漂浮垃圾分类检测技术研究[J/OL]. 金属矿山: 1-11. [2021-06-20]. http://kns. cnki. net/kcms/detail/34.1055. TD.20210608.1117.005.html.

LI D, YAN Z, SUN J. Study on classification and de-

tection technology of river floating garbage based on UAV vision[J/OL]. Metal Mine: 1-11. [2021-06-20]. http://kns. cnki. net/kcms/detail/34.1055. TD. 20210608. 1117.005.html.

[25] SIFRE L, MALLAT S. Rigid-motion scattering for texture classification[J]. Computer Science, 2014, 3559: 501-515.

Yield Estimation Method of Apple Tree Based on Improved Lightweight YOLOv5

LI Zhijun[1,2], YANG Shenghui[1,2], SHI Deshuai[1,2], LIU Xingxing[1,2], ZHENG Yongjun[1,2*]

（1. *College of Engineering*, *China Agricultural University*, *Beijing* 100083, *China*; 2. *Yantai Institute of China Agricultural University*, *Yantai* 264670,*China*）

Abstract: Yield estimation of fruit tree is one of the important works in orchard management. In order to improve the accuracy of in-situ yield estimation of apple trees in orchard, a method for the yield estimation of single apple tree, which includes an improved YOLOv5 fruit detection network and a yield fitting network was proposed. The in-situ images of the apples without bags at different periods were acquired by using an unmanned aerial vehicle and Raspberry Pi camera, formed an image sample data set. For dealing with no attention preference and the parameter redundancy in feature extraction, the YOLOv5 network was improved by two approaches: 1) replacing the depth separable convolution, and 2) adding the attention mechanism module, so that the computation cost was decreased. Based on the improvement, the quantity of fruit was estimated and the total area of the bounding box of apples were respectively obtained as output. Then, these results were used as the input of the yield fitting network and actual yields were applied as the output to train the yield fitting network. The final model of fruit tree production estimation was obtained by combining the improved YOLOv5 network and the yield fitting network. Yield estimation experimental results showed that the improved YOLOv5 fruit detection algorithm could improve the recognition accuracy and the degree of lightweight. Compared with the previous algorithm, the detection speed of the algorithm proposed in this research was increased by up to 15.37%, while the mean of average accuracy (*mAP*) was raised up to 96.79%. The test results based on different data sets showed that the lighting conditions, coloring time and with white cloth in background had a certain impact on the accuracy of the algorithm. In addition, the yield fitting network performed better on predicting the yield of apple trees. The coefficients of determination in the training set and test set were respectively 0.7967 and 0.7982. The prediction accuracy of different yield samples was generally stable. Meanwhile, in terms of the with/without of white cloth in background, the range of relative error of the fruit tree yield measurement model was respectively within 7% and 13%. The yield estimation method of apple tree based on improved lightweight YOLOv5 had good accuracy and effectiveness, which could achieve yield estimation of apples in the natural environment, and would provide a technical reference for intelligent agricultural equipment in modern orchard environment.

Key words: apple in-situ yield estimation; deep learning; fruit detection; BP neural network; YOLOv5